

Methoden und Probleme der Kontrolle sozialer Erwünschtheit (Social Desirability)*

Hans Dieter Mummendey Universität Bielefeld

Zusammenfassung

Das psychologische Konzept der sozialen Erwünschtheit (social desirability response set) wird grundsätzlich als differenzierungsbedürftig angesehen. Es werden Methoden der Kontrolle sozialer Erwünschtheitstendenzen (durch Itemkonstruktion, durch Antwortkombination, durch Kontrollskalen, durch Instruktion) behandelt, und es wird als Verfahren der experimentellen Kontrolle sozialer Erwünschtheit durch Manipulation der Rahmensituation das Bogus-Pipeline-Verfahren besprochen. Da sowohl bei Reaktionen im Sinne sozialer Erwünschtheit als auch bei Versuchen zu ihrer Kontrolle Täuschungsprozesse eine Rolle spielen, wird auf ethische Probleme der Kontrolle sozialer Erwünschtheit eingegangen.

Abstract

Social desirability is regarded as a psychological construct that has to be differentiated into several partial constructs. A review of methods of control for social desirability response set is given, stressing the bogus pipeline paradigm as an efficient strategy to modify the total response situation. Since social desirability reactions and all methods to control them make use of deception, several ethical implications of controlling social desirability are discussed.

1. Das Konzept sozial erwünschter Reaktionen

„Sozial erwünscht“ (social desirable) zu reagieren heißt entsprechend der Erwartung zu reagieren, daß die Art der Reaktion die Zustimmung einer sozialen Gruppe findet. Schon früh wurde erkannt (z. B. durch CRONBACH, 1946, 1950; EDWARDS, 1953, 1957a), daß verbale Reaktionen auf verbale Stimuli, also Feststellungen oder Fragen, bezüglich ihrer Bejahungswahrschein-

lichkeit unter anderem von dem Grad abhängen, mit dem die Bejahung als sozial erwünscht beurteilt wird. Damit wird ein psychologisches Konstrukt etabliert bzw. eine psychologische Variable „soziale Erwünschtheit“ (social desirability) konstruiert, die als wichtige Fehlerquelle für sämtliche als „subjektive“ Verfahren klassifizierbaren psychologischen Meßmethoden fungiert.

Das *Schwellen-Modell* der Persönlichkeitsmessung von VOYCE & JACKSON (1977) mag zeigen, wie man sich die wichtigsten unterschiedenen klassischen Reaktionstendenzen (response sets) als Spielarten der Ten-

* Veränderte Fassung des Positionsreferates auf dem 32. Kongreß der Deutschen Gesellschaft für Psychologie in Zürich 1980, „Soziale Erwünschtheit als Problem psychologischer Forschung“.

denz, sozial erwünscht zu reagieren, vorstellen kann. VOICE & JACKSON beschreiben diese „stilistischen“ (im Unterschied zu „inhaltlichen“) Reaktionen von Versuchspersonen (Vpn) auf Items in Persönlichkeitsfragebogen als abhängig von zwei Prozessen: der Sensibilität der Vp gegenüber dem Erwünschtheitsgrad eines Items (d. h. dem Social Desirability Scale Value SDSV; vgl. EDWARDS, 1957 a) und der Schwelle, im sozial erwünschten Sinne zu antworten. (Diese beiden Prozesse entsprechen den konventionellerweise als „Erwünschtheits-“ und „Jasagetendenz“ bezeichneten Response Sets bzw. Styles.) Das Modell geht davon aus, daß die Wahrscheinlichkeit einer positiven Reaktion auf ein Item als Funktion der Erwünschtheit des Items variiert. Betrachtet man dabei individuelle Unterschiede, so nimmt das Modell an, daß sich für jedes Individuum eine Kurve zeichnen läßt, die die beurteilte Item-Erwünschtheit und die Bejahung des Items durch die Person miteinander in Beziehung setzt.

Diese Kurve ist von zwei Parametern bestimmt: a) der Schwelle, d. h. dem kritischen Grad an Erwünschtheit, der den Übergang von einer „Falsch“- zu einer „Richtig“-Reaktionstendenz markiert, und b) dem Salience-Parameter, der eine Schätzung der Sensitivität der Person gegenüber der Erwünschtheitsdimension gibt. Jede Person handelt so, als „kenne“ sie den Erwünschtheitsgrad jedes Items. Liegt ein Item unterhalb ihres eigenen Schwellenwerts, so wird die Person dieses Item nicht bejahen. Liegt der Erwünschtheitswert des Items über ihrer Schwelle, so wird sie das Item bejahen. Eine Person mit einer niedrigen Schwelle wird voraussichtlich eine größere Zahl von „Stimmt“-Reaktionen geben, während eine Person mit einem hohen Schwellenwert eher „Stimmt nicht“-Antworten geben wird, so daß sich dadurch *individuelle Differenzen in der Jasagetendenz für Items im neutralen Bereich der Erwünschtheit* ergeben.

Da es in Differentieller und Diagnostischer Psychologie erklärtermaßen um *individuell* zurechenbare Reaktionen – unbeeinflusst von dem, was *sozial* erwünscht sein mag – geht, kommt der Variablen „Soziale Erwünschtheit“ der Status einer die Validität psychologischer Messungen mindernden *Suppressorvariablen* zu (vgl. HORST, 1941; DICKEN, 1963; WIGGINS, 1973), also einer mit einem Prädiktor, nicht jedoch dem entsprechenden Kriterium korrelierenden Variablen.

Der Suppressor-Charakter der menschlichen Tendenz, im sozial erwünschten Sinne zu reagieren, erscheint dadurch plausibel, daß jede verbale Äußerung bzw. jede Antwort einer Vp nicht nur hinsichtlich ihres „absoluten Inhalts“, sondern ebenfalls hinsichtlich des Grades wahrge-

nommener bzw. erwarteter sozialer Erwünschtheit skaliert werden kann. Es ist aus diesem Grunde allerdings sehr schwierig, „Inhalt“ und Anteil der „Erwünschtheit“ voneinander zu trennen.

Ein Statement, dessen Social-Desirability-(SD-)Charakter den Mitgliedern unserer Gesellschaft unmittelbar einleuchten wird, ist das von MORF & JACKSON (1972) apostrophierte „I usually help old ladies across the street“. Bei dieser Feststellung handelt es sich um eine Beschreibung offenen *Verhaltens*, und eine Bejahung dieser Verhaltensweise aus dem Grunde, daß die Vp tatsächlich bei jeder sich bietenden Gelegenheit älteren Damen über die Straße hilft (ausgenommen vielleicht, sie sei selbst alt und gebrechlich, was bei den zumeist untersuchten Vpn aber nicht der Fall ist), ist höchstens so plausibel wie die allgemeine Neigung der Vp, Testfragen eher mit „Stimmt“ zu beantworten, oder aber das Statement zu bejahen, weil das somit sich selbst zugeschriebene Verhalten einer anerkannten sozialen Norm entspricht. Das gleiche gilt für alle Bejahungen von Statements, die im Sinne von EDWARDS (1957 b) echte *Einstellungs-Statements* sind, insofern sie sich nicht auf Fakten, sondern auf die Evaluation von Objekten beziehen. Die Bejahung z. B. der Feststellung „Türkische Gastarbeiter sollten in jeder Hinsicht wie Mitbürger behandelt werden“ läßt sich prinzipiell in gleicher Weise mehrdeutig interpretieren wie das erwähnte Verhaltens-Statement, und eine der mehreren möglichen Interpretationen der Bejahung wird diejenige im Sinne des SD-Konstrukts sein.

Während sich Vorschläge zur Kontrolle der allgemeinen *Jasagetendenz* aufgrund des vollständig „formalen“ Charakters dieses Konzepts grundsätzlich erfolgreicher realisieren lassen, z. B. weil es in Grenzen die Möglichkeit gibt, sprachlich formulierte Sachverhalte in Abschattungen positiv oder negativ auszudrücken und auf diese Weise Listen von Statements herzustellen, die hinsichtlich des Grades an positiver bzw. negativer Formulierung bzw. Verschlüsselung abgestuft bzw. ausbalanciert sind und so eine indirekte Kontrolle der formalen *Jasagetendenz* ermöglichen (vgl. JACKSON, 1966; MESSICK,

1966; KEIL, 1968, 1973; BENTLER, JACKSON & MESSICK, 1971; EHLERS, 1973), scheint das Problem der Nicht-Separierbarkeit von „sozialer Erwünschtheit“ und Inhalten deswegen schwieriger zu sein, weil es sich bei sozialer Erwünschtheit stärker um eine „inhaltliche“ als um eine „formale“ Störvariable zu handeln scheint. Der Grad der „Positivität“ eines Fragebogenitems beispielsweise kann variiert werden, ohne daß der skalierbare Gehalt des Items allzusehr verändert wird; mit der Manipulation des sozialen Erwünschtheitsgrades eines Items wird dagegen in der Regel eine stärkere Veränderung des Item-Inhaltes vorgenommen.

Auf dem Fragebogen-Symposium des Kieler Kongresses der Deutschen Gesellschaft für Psychologie 1970 wies JANKE (1973) darauf hin, „daß fast alle Items in unseren Fragebogen Verhaltensmerkmale erfragen, die in unserer Gesellschaft sich hoher oder geringer Wertschätzung erfreuen, d.h. sozial erwünscht oder unerwünscht sind. Neutrale Items sind in der Regel schwach repräsentiert“ (p. 46). Dies erscheint noch als vorsichtig formuliert. Der MMPI als Vater (oder Großvater) der meisten gegenwärtigen Persönlichkeitsfragebogen enthält kaum ein Item, das als *neutral* in bezug auf seine Wertung zu bezeichnen wäre. Schon BLOCK (1965) konnte die beiden ersten aus den Items des MMPI extrahierten Faktoren als Social Desirability und Acquiescence interpretieren. Eigene Versuche, von unabhängigen Beurteilern als „neutral“ (im Gegensatz zu erwünscht bzw. unerwünscht) zu bezeichnende Statements für eine SD-Kontrolluntersuchung zu finden, sind kürzlich völlig gescheitert. Die zumeist hohen Korrelationen zwischen Maßen sozialer Erwünschtheit und Persönlichkeitsmerkmalen, die auf der Basis von Skalen bzw. Itemlisten immer wieder gewonnen werden, lassen sich auf hohe Korrelationen zwischen sozialer Erwünschtheit und Inhalt auf Item-Ebene zurückführen. Items mit extrem hohen SD-Werten zeigen in der Regel hohe Bejahungsfrequenzen, und Items mit extrem negativen SD-Werten zeigen hohe Verneinungsfrequenzen; hierbei dominieren Korrelationen um bzw. über 0.90 (vgl. ANGLEITNER, 1976). Auch die vielen Versuche, Adjektive zur Beschreibung

von Persönlichkeitseigenschaften hinsichtlich sozialer Erwünschtheit beurteilen zu lassen (z. B. durch ANDERSON, 1968; COWEN & BUDIN, 1964; SCHÖNBACH, 1972; KLAPPROTT, 1972; BUXBAUM, 1976, 1979) ergeben stets zweigipflige Häufigkeitsverteilungen, wenn man die Gesamtmenge der beurteilten Persönlichkeitseigenschaften betrachtet (vgl. *Abbildung 1*). Diese Likableness-Ratings von Eigenschaften korrelieren, wie KIRBY & GARDNER (1971) zeigten, 0.98 mit den EDWARDS'schen SDSV, d.h. den Erwünscht-unerwünscht-Ratings derselben Eigenschaften. Betrachtet man die SDSV von fast 3000 Statemenes aus Persönlichkeitsfragebogen, so verteilen sich auch diese klar zweigipflig (EDWARDS, 1970). Demnach scheint es fast unmöglich zu sein, verbale Beschreibungen von

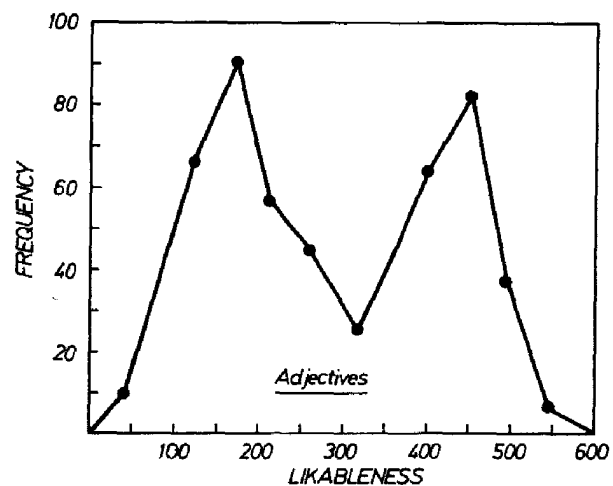


Abbildung 1:
Häufigkeitsverteilung der Erwünschtheits-Ratings von 555 Persönlichkeitseigenschaften (ANDERSON, 1968)

Verhaltensweisen oder Einstellungen von Mitmenschen oder sich selbst zu erhalten, die hinsichtlich sozialer Evaluation neutral sind. Es ist WIGGINS (1979) zuzustimmen, der Persönlichkeitsvariablen, die nicht hinsichtlich Erwünschtheit variieren, als armselige, schwache Repräsentationen von Kategorien des wirklichen Lebens ansehen möchte.

Der Suppressoreffekt der SD-Tendenz wurde selten empirisch exakt bestimmt; nur so erscheint es auch verständlich, wenn am Ende des sehr umsichtigen Berichtes von EHLERS (1973)

über die Effektivität der Kontrollen von Reaktionseinstellungen resümiert wird, daß der Einfluß von Reaktionseinstellungen auf die Gültigkeit von Fragebogen vermutlich überschätzt werde und daß eine Berücksichtigung entsprechender Konstrukte im Sinne intervenierender Variablen nicht lohnenswert sei. Bei diesem Urteil wird allerdings (zu Recht) vor allem dem zu erwartenden Kontroll-Aufwand besonderes Gewicht beigemessen (vgl. dazu die nächsten Abschnitte). Die Untersuchung von DICKEN (1963) auf der Basis des CPI hatte zwar nur in vier von 24 bzw. in zwei von 36 verglichenen Fällen einen signifikanten Suppressoreffekt ergeben, und diejenige von GOLDBERG, RORER, GREENE (1970) auf der Basis des gleichen Tests hatte nur bei 30 von 169 Vergleichen einen stabilen Suppressoreffekt aufgewiesen. Dagegen zeigen aber viele Untersuchungen mit manipulierter Instruktion (z. B. Anweisungen an die Vpn, sich möglichst günstig darzustellen) enorme Beschönigungstendenzen noch im Vergleich zu völlig unbeeinflussten Kontrollgruppen, bei denen vermutlich bereits SD-Tendenzen wirksam sind, auf. So fanden u. a. HÄCKER, SCHWENKMEZGER & UTZ (1977) bei ca. der Hälfte der Skalen verschiedener Fragebogentests signifikante Verfälschungseffekte; sie schlossen daraus unter anderem, daß ein Teil der von CATTELL als unverfälschbar gekennzeichneten Objektiven Tests ihrem Anspruch auf Invarianz gegenüber Motivationseffekten nicht gerecht werde. Die praktischen Erfahrungen von Diagnostikern, die beispielsweise Tests vom Q-Typus in ausleseähnlichen Situationen wie z. B. beim Technischen Überwachungsverein oder bei Strafgefangenen, kurz: in sogenannten Ernstsituationen angewendet haben (vgl. z. B. BÖCHER, 1967; SCHMIDT & VORTHMANN, 1971; REITZNER, 1974; HÄCKER et al., 1977) sind sehr überzeugend – es zeigen sich erhebliche Beschönigungseffekte, wenn man die Daten mit in solchen Situationen gewonnenen vergleicht, in denen keine forcierte Beschönigungsabsicht zu erwarten ist.

Es wird deutlich, daß das psychologische Konzept sozial erwünschter Reaktionen prinzipiell *Täuschungsverhalten seitens der untersuchten Person* impliziert. Je nach Situation, Thema und

Persönlichkeit der Vp wird es zu mehr oder weniger starken Substitutionen von als „wahr“ oder „echt“ aufgefaßten durch im Sinne sozialer Normen korrigierte bzw. verfälschte Reaktionen kommen. Weiter unten wird zu diskutieren sein, wie der Psychologe mit eigenen Mitteln der Täuschung versucht, das aus seiner Forschungsperspektive gestörte Gleichgewicht zwischen Versuchsleiter und Versuchsperson wiederherzustellen.

2. Differenzierungen des Erwünschtheits-Konzepts

Mit den zuletzt angestellten Überlegungen ist bereits angesprochen, daß die Tendenz, sozial erwünscht zu reagieren, aus unterschiedlichen Perspektiven gesehen werden kann, wie es die Begriffe „Reaktionstendenz“, „Fehler“, „Bias“ etc. mit ihren jeweils unterschiedlichen Konnotationen andeuten. Hier soll allerdings nicht der Platz für begriffliche Spitzfindigkeiten sein. Dennoch muß auf die schon früh (vgl. JACKSON & MESSICK, 1958) erfolgte Unterscheidung zwischen *Response-Set* und *Response-Style* kurz eingegangen werden, obgleich sie nicht immer einheitlich ist (vgl. RORER, 1965). Autoren, die diese Unterscheidung machen, heben zumeist auf den Grad der Situationsspezifität bzw. Generalität der SD-Tendenz ab und fassen soziale Erwünschtheit als „Stil“ im Sinne eines verschiedenen Antwortsituationen übersteigenden Persönlichkeitsmerkmals auf (vgl. JACKSON & MESSICK, 1958; BENTLER, JACKSON & MESSICK, 1971). Wenn gelegentlich gezeigt wurde, daß Personen mit hohen SD-Scores beispielsweise

- besser konditionierbar sind (CROWNE & MARLOWE, 1964; EPSTEIN, 1964)
- diverse Lernaufgaben besser erledigen (BROWN, 1960; NORMAN, 1963; STEPHENS, zit. n. ALLAMAN, JOYCE & CRANDALL, 1972; KATZ, HENCHY & ALLEN, 1968)
- motorische Aufgaben schneller ausführen können (FORD & RUBIN, 1970),

aber auch

- beeinflussbarer bzw. konformer sind (CROWNE & MARLOWE, 1964)
- weniger Self-Esteem besitzen und unsicherer sind (CROWNE & MARLOWE, 1964)
- Mütter mit relativ höheren Ausprägungsgraden an Feindseligkeit, Kritizismus, Restriktivität, Zwang und Bestrafungstendenz bei Mädchen und Jungen, bei Jungen zusätzlich mit weniger Ermutigung zur Entwicklung von Fertigkeiten als Erziehungsverhaltensweisen gehabt haben (ALLAMAN, JOYCE & CRANDALL, 1972),

so wird hieraus die radikale Konzeption sozialer Erwünschtheitstendenz als eine *Persönlichkeits-eigenschaft* deutlich. Allerdings hat sich wiederholt gezeigt, daß die Generalität der SD-Tendenz bereits über unterschiedliche Meßmethoden hinweg recht begrenzt ist (vgl. FISKE & PEARSON, 1970; EDWARDS & ABBOTT, 1973). Es erscheint daher als eine Frage der persönlichen Toleranz des Forschers gegenüber den sogenannten neuen Erkenntnissen der in den 70er Jahren aufgeflackerten Person-Situation-Interaktions-Kontroverse (vgl. MISCHEL, 1968; ENDLER & MAGNUSSON, 1976; MAGNUSSON & ENDLER, 1977; HERRMANN, 1980), ob bzw. in welchem Maße die SD-Tendenz als Stil bzw. Quasi-Persönlichkeitsmerkmal formaler Art aufrechterhalten ist.

Die Annahme oder Aufdeckung großer Varianzanteile in psychologischen Messungen, die offensichtlich stärker auf die Orientierung an sozialen Normen als auf eine verschärfte Wahrnehmung und Beurteilung eigener Einstellungen und Verhaltensweisen zurückführbar sind, hat die meisten Response-Set-Forscher dazu geführt, in relativ undifferenzierter Weise ein *allgemeines* Konstrukt „soziale Erwünschtheit“ anzunehmen, sei dieses nun als Set oder Style mehr oder weniger situationsspezifisch. Betrachtet man die Vielzahl der schon erwähnten SD-Ratings von Items und Adjektiven hinsichtlich ihrer sozialen Erwünschtheit mit dem Ziel der Zuschreibung von SD-Indizes, oder die vielen Versuche mit Instruktionsmanipulationen, denen gemäß sich die Vp vorübergehend in ein positives Licht setzen soll (vgl. weiter unten), so

fällt auf, daß stets von der Konzeption eines allgemeinen, übersituativ und gruppen-unspezifisch existierenden Standards für sozial positive Bewertungen bzw. soziale Wünschbarkeit ausgegangen wird. Dieses Vorgehen erscheint dadurch gerechtfertigt, daß die SDSV von Items gewöhnlich über verschiedene Gruppen von Beurteilern hinweg hoch korrelieren (EDWARDS, 1970). Andererseits hat eine Reihe von Untersuchungen ergeben, daß Bewertungen von Verhaltensweisen in Abhängigkeit von *Alter*, *Geschlecht* und *sozialer Schicht* variieren und daß es zu Wechselwirkungen zwischen diesen Merkmalen bezüglich der Erwünschtheit von Verhaltensweisen und Eigenschaften kommt (vgl. z. B. CROTT & ROSSRUCKER, 1974). Soziale Erwünschtheitsurteile fallen auch unterschiedlich aus je nachdem, ob *Selbsteinschätzungen* oder ob *Fremd-Selbsteinschätzungen* abgegeben werden (vgl. CROTT, PRÜFER & WOLFSHÖRNDL, 1977). *Altersunterschiede* in Erwünschtheits-Ratings wurden ferner von MILTURN & LEWIS (1968), AHAMMER (1971) sowie AHAMMER & BALTES (1972) gefunden; sie scheinen besser gesichert als beispielsweise *Geschlechts-* und *Bildungsunterschiede* (vgl. z. B. KLETT, 1957; EDWARDS, 1957a; MARLOWE & CROWNE, 1962; ALLAMAN, JOYCE & CRANDALL, 1972; KLEIN, 1974; abweichend z. B. CRANDALL, CRANDALL & KATKOVSKY, 1965, für das *Kindesalter*). Mit großer Konsistenz wurden dagegen *interkulturelle* Übereinstimmungen hinsichtlich Beurteilungen nach sozialer Erwünschtheit erzielt: Norwegische, arabische, niseische und amerikanische Vpn beurteilten den SD-Gehalt von EPPS-Items nicht wesentlich unterschiedlich (EDWARDS, 1957a; LOVAAS, 1958; KLETT & YAUKEY, 1959), während TRIANDIS (1964) Entsprechendes für Griechen und Amerikaner zeigte. In anderen Fällen (vgl. z. B. KLEIN, 1974) ergaben sich jedoch Nationalitätenunterschiede hinsichtlich SD-Ratings von Eigenschaften, z. B. beim Vergleich von Deutschen und Franzosen.

Allein die Tatsache der Altersspezifität von Erwünschtheits-Urteilen, wie sie mehrfach übereinstimmend gefunden wurde, erscheint ausreichend, um die Annahme eines allgemeinen, gruppen-unspezifischen SD-Konstrukts zurück-

zuweisen. Ergänzt wird dieses Bild durch die gelegentlich gefundenen niedrigen Korrelationen zwischen verschiedenen Kontrollskalen sozialer Erwünschtheit, die in der Regel aus Feststellungen mit scheinbar allgemein hohem Erwünschtheitsgrad bestehen (vgl. EHLERS, 1973). Es wird deutlich, daß eine *generelle* Konzeption sozialer Erwünschtheit, und sei es auch nur für die Beantwortung von Persönlichkeitsfragebogen als einer bestimmten *Methode* – jede Untersuchungsmethode konstituiert ja zugleich eine spezifische *Situation* – problematisch ist und daß man wohl kaum umhinkommt, auf *spezifische* Stichproben von Methoden und *Personen*, Situationen und *Themen* bezogene *Erwünschtheitskonzepte* zu formulieren. Wenn es auch unökonomisch erscheint, in der letzten Konsequenz für jede empirische Untersuchung ein eigenes Erwünschtheitskonzept zu etablieren und zu operationalisieren, so erscheint doch wenigstens der Anspruch berechtigt, für psychologische Messungen *Teilkonzepte sozialer Erwünschtheit* zu formulieren, sofern in einem bestimmten Bereich – wie z. B. für unterschiedliche Lebensalter und Kohorten – bereits gesicherte Erkenntnisse vorliegen.

Im äußersten Falle wäre das Konzept sozialer Erwünschtheit durch dasjenige einer „*persönlichen* Erwünschtheit“ zu ersetzen. Schon SCOTT

(1963) hielt aufgrund von Untersuchungen mit dem EPPS soziale Erwünschtheit nicht für ein Item-Charakteristikum, sondern für das Ergebnis einer *Item-Person-Interaktion*. Er gibt zugleich den auch für eine auf lernpsychologischen Grundlagen betriebene Selbstkonzeptforschung interessanten Hinweis, daß *die Erwünschtheit eines Items* auch die *Konsequenz der Bejahung* sein könne, und nicht nur umgekehrt. Diese Interpretation des Verhältnisses von Inhalt und Set bzw. Style ist – vielleicht mangels lerntheoretischer Orientierung der meisten differentiell- und sozialpsychologischen Autoren – nicht weiter beachtet worden. Schon 1953 hatte EDWARDS gezeigt, daß SD-Items zugleich solche sind, die man sich zuschreibt bzw. daß Merkmale, die man sich zuschreibt, auch hochgradig „sozial erwünscht“ sind. NANCY WIGGINS (1966) zeigte, daß die MMPI-Items sechs funktional unabhängigen Ideal-Personen, d. h. hypothetischen Individuen mit besonders erwünschten Eigenschaften zuzuordnen sind; sie reflektieren sehr unterschiedliche Cluster von Items, jeweils mit hohen SDSV. Aufgrund der Anwendung des Verfahrens von TUCKER & MESSICK (1963) sieht die Autorin die Multidimensionalität des SD-Konzepts als erwiesen an und folgert daraus differentielle Erwünschtheitsstrukturen für Items *und* Individuen.

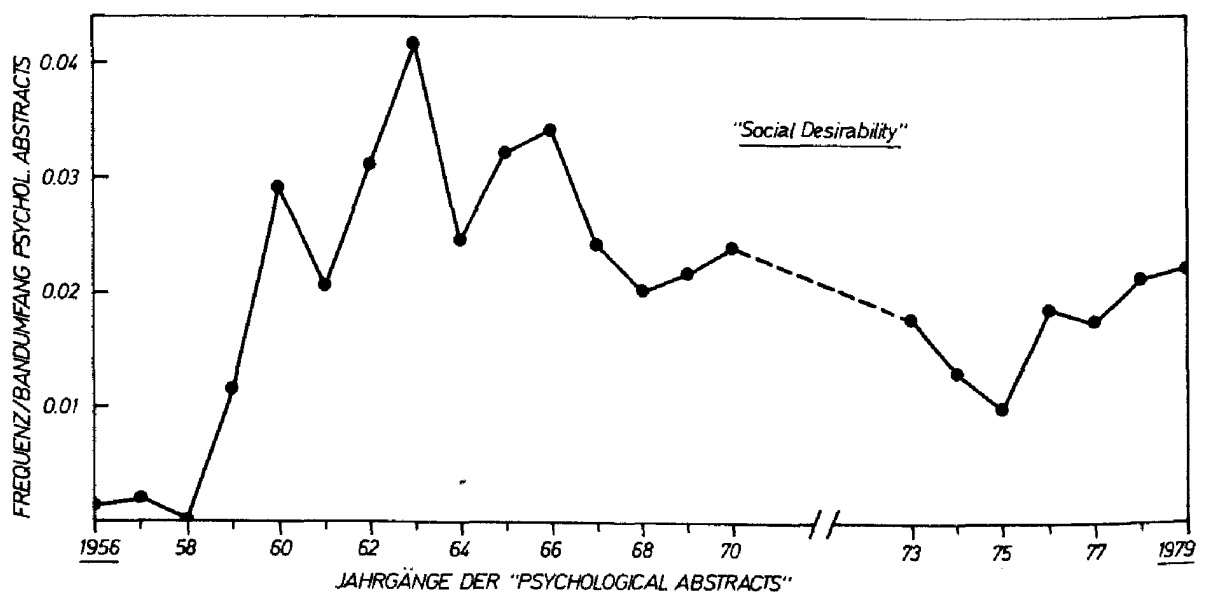


Abbildung 2: Relative Häufigkeit von Publikationen zum Stichwort „Social Desirability“ in den „Psychological Abstracts“ der Jahrgänge 1956 bis 1979.

Damit scheinen genügend Hinweise auf die Notwendigkeit einer *Differenzierung des Konzepts „Soziale Erwünschtheit“* vorzuliegen. Für die Praxis der psychologischen Forschung ist jedoch aufgrund einer Durchsicht der Literatur nicht etwa festzustellen, daß die notwendigen Arbeiten an Teilkonstrukten sozialer Erwünschtheit in den 70er Jahren stattgefunden hätten. Vielmehr ist ein relativer Rückgang an Publikationen über soziale Erwünschtheit festzustellen, wie *Abbildung 2* zeigt. Die absolute Häufigkeit der Veröffentlichungen (in *Abb. 2* jedoch bezogen auf die Gesamtmasse an Publikationen, die in den *Psychological Abstracts* referiert werden) zum Stichwort „Social Desirability“ wird obendrein vor allem dadurch genährt, daß SD-Messungen bei den verschiedensten empirischen Untersuchungen, z. B. klinisch-psychologischer Art, „mitlaufen“. Statt einer hier wie überall anderswo in unserem Fach erforderlichen Differenzierung beobachten wir also womöglich ein langsames Fallenlassen des Themas. Auf das Konstrukt „Soziale Erwünschtheit“ – gewissermaßen als Allerweltskonstrukt, als *deus ex machina* – wird überall dann zurückgegriffen, wenn es an die Interpretation hypothesennonkonformer Ergebnisse empirischer Untersuchungen geht. Den SD-Tendenzen geht es damit ähnlich wie den verwandten Demand Characteristics, Evaluation Apprehensions und weiteren Vpn-Effekten der Sozialpsychologie des Experiments (vgl. z. B. ORNE, 1962; PAGE, 1973) – sie sind für globale Vermutungen und schnelle Erklärungen gut, werden aber selten in ihrer spezifischen Wirkung nachgewiesen.

3. Die Kontrolle sozialer Erwünschtheits-Tendenzen

Bemühungen um Kontrolle von SD-Tendenzen in Persönlichkeitsmessung und Diagnostik gibt es seit den ersten Beschreibungen des SD-Konstrukts. Die verschiedenen Kontrollversuche lassen sich wie folgt gliedern:

- Kontrolle durch *Itemkonstruktion*
- Kontrolle durch *Antwortkombination*
- Kontrolle durch spezielle *Kontrollskalen*
- Kontrolle durch *Instruktion*.

Über eine Reihe von Ergebnissen solcher Kontrollversuche wurde bereits – insbesondere von EHLERS (1973) – in dem Fragebogen-Symposium auf dem 27. Kongreß der Deutschen Gesellschaft für Psychologie in Kiel 1970 berichtet. Da EHLERS zusammenfassend den Einfluß von Reaktionstendenzen auf die Fragebogengültigkeit als überschätzt ansieht („Eine Berücksichtigung von Reaktionseinstellungskonstrukten im Sinne intervenierender Persönlichkeitsmerkmale lohnt sich nicht“; 1973, p. 51) sei der Vollständigkeit halber zusätzlich auf die grundsätzliche Möglichkeit der *Nichtkontrolle* als einer sinnvollen Coping-Strategie hingewiesen. Im folgenden sollen die unterschiedlichen Kontrollmöglichkeiten nur kurz besprochen werden.

3.1 Kontrolle durch Itemkonstruktion

Insofern die Durchschaubarkeit dessen, was mit einem psychologischen Verfahren gemessen werden soll, die absichtliche oder unabsichtliche Reaktion der Vp, im sozial erwünschten Sinne zu antworten, fördern kann, kann versucht werden, bei der *Itemselektion* solche Items zu präferieren, die hinsichtlich des zu messenden Merkmals möglichst wenig *transparent* sind. Solche Kontrollversuche stoßen jedoch bei Selbstberichten über Verhalten ebenso wie bei der Einstellungsmessung auf grundsätzliche Schwierigkeiten. Inhalt und Set bzw. Style lassen sich nämlich nicht dadurch trennen, daß man gewissermaßen den Inhalt auszuschalten versucht. Auf der anderen Seite mag es durchaus Statements geben, die relativ wenig SD-behaftet sind und dennoch ausreichende Trennschärfe besitzen. Beim Versuch der SD-Kontrolle auf Item-Ebene geht es somit darum, „Skalen zu konstruieren, die verglichen mit ihrer inhaltlichen Trait-Varianz einen relativ höheren Anteil an Trait- als an sozialer Erwünschtheitsvarianz aufweisen“ (ANGLEITNER, 1976, p. 92). Dies ist offensichtlich mit Items, die direkt oder indirekt

dem MMPI entstammen, nicht möglich; vielfach zeigte sich (vgl. z. B. SCHMIDT & VORTHMANN, 1971), daß die meisten SD-Items identisch mit Extraversion- und Neurotizismus-Items sind; umgekehrt sind z. B. die Items der MAS-Skala (TAYLOR, 1953) zu mehr als der Hälfte identisch mit der EDWARDSSchen SD-Kontrollskala, die gleichfalls aus dem MMPI stammt (ANGLEITNER, 1976).

Ein Versuch der SD-Kontrolle bei der Itemkonstruktion wurde in bislang systematischster Weise bei der Konstruktion des PRF von JACKSON (1974) vorgenommen. Um Statements mit relativ hoher SD-Varianz bei der Selektion zu eliminieren, wurden Korrelationen mit einem gleichzeitig erhobenen SD-Score ermittelt und mit der Item-Trennschärfe verglichen. Aus der Differenz der Determinationskoeffizienten Item/Test und Item/SD gewann JACKSON dann einen „differentiellen Reliabilitäts-Index“, der als Kriterium für die Itemselektion diente. Im Grunde handelt es sich dabei um die Anwendung des Prinzips spezieller SD-Kontrollskalen (vgl. weiter unten) bereits bei der Itemselektion. Das Verfahren wurde auch bei der Konstruktion einer deutschen Version des PRF angewendet (ANGLEITNER, STUMPF & WIECK, 1976).

Eine mögliche Schwierigkeit dieses Kontrollverfahrens könnte darin bestehen, daß sich in einer Menge solcher selektierten Items, die aufgrund der Vorermittlung als relativ wenig SD-anfällig anzusehen sind, neue Erwünschtheits-Normen herausbilden. Da die Itemselektion letztlich auf das Kriterium der Korrelation mit extremen SD-Items zurückgeht, könnten sich in veränderter Itemkombination neue Erwünschtheits-Hierarchien unter den Items ergeben, da Inhalt und Set nur relativ verschoben, nicht aber absolut getrennt werden können, was ihren relativen Varianzbeitrag betrifft.

3.2 Kontrolle durch Antwortkombination

Da soziale Erwünschtheit und Bejahung direkt miteinander zusammenhängen – je sozial erwünschter ein Inhalt, desto eher wird ihm zuge-

stimmt – können die der Vp angebotenen Antwortalternativen hinsichtlich des Grades ihrer sozialen Erwünschtheit ausbalanciert werden. Systematisch geschieht dies in der *Forced-Choice-Technik*, wie sie beispielsweise im EPPS von EDWARDS (1954) realisiert ist. Das Prinzip balancierter Antwortalternativen läßt sich insbesondere bei Persönlichkeitsinventaren, die mehr als ein einziges Persönlichkeitsmerkmal messen möchten, anwenden, da dann der Vp stets beispielsweise Paare unterschiedlich valider, aber gleichermaßen erwünschter Statements vorgegeben werden können.

Abgesehen von der relativen Unökonomie der Forced-Choice-Technik läßt sich u. U. gegen die Methode auch einwenden, daß es nach vollzogener Selektion der Antwortkombinationen zu einer Reorganisation der SD-Hierarchie der Items kommen kann (vgl. auch COWEN, BUDIN & BUDIN, 1964). Gegen den Paarvergleich von SD-balancierten Items spricht jedoch auch grundsätzlich die mögliche Mehrdimensionalität des Social-Desirability-Konzepts (vgl. SCOTT, 1963; N. WIGGINS, 1966), die eine Ausbalancierung von Items gemäß einem allgemeinen, „mittleren“ Erwünschtheitswert als unzweckmäßig, ja sinnlos erscheinen ließe. In letzter Konsequenz wären entsprechende Paarungen stets nur für ein bestimmtes Individuum ausbalanciert. Die empirisch unterstützten Argumente von SCOTT und WIGGINS richten sich allerdings nicht nur gegen die Forced-Choice-Kontrolle, sondern lassen alle aufgeführten Kontrollverfahren mit Ausnahme derjenigen durch Instruktionsmanipulation als vergeblich erscheinen.

3.3 Kontrolle durch spezielle Kontrollskalen

Diese am häufigsten angewendete Methode zur Bestimmung und Reduktion von SD-Varianz soll hier nur kurz erwähnt werden. Die seit den 50er Jahren existierenden SD-Kontrollskalen (z. B. EDWARDS, 1957a; CROWNE & MARLOWE, 1960; deutschsprachig z. B. LÜCK & TIMAEUS, 1969; SCHMIDT & VORTHMANN, 1971; DICKENBERGER, HOLTZ & GNECH, 1978) dienen zu meist der statistischen Kontrolle durch Auspar-

tialisierung der SD-Variablen im nachhinein. Sie können nach den Regeln der klassischen Testkonstruktion erstellt werden, indem entweder die SDSV von Statements durch Rater bestimmt oder aus Untersuchungen mit manipulierter Instruktion (vgl. weiter unten) hervorgegangene, besonders SD-anfällige Statements ausgelesen worden sind. Das in die diversen Kontrollskalen gesetzte Vertrauen muß aber schwinden, wenn sich häufig herausstellt, daß sie untereinander nur in geringem Maße korrelieren (vgl. EHLERS, 1973; ANGLEITNER, 1976) und – wie bereits erwähnt – mit einer ganzen Reihe von Persönlichkeitsvariablen kovariieren, so daß „die verschiedenen sozialen Erwünschtheitsmaße häufig höher mit Persönlichkeitsmaßen als untereinander korrelieren“ (ANGLEITNER, 1976, p. 91). Müßte in der Konsequenz „eine auf die jeweilige Stichprobe bezogene Itemanalyse“ vorgenommen werden, so erwiesen sich die SD-Kontrollskalen in der Tat nur „scheinbar“ als ökonomischstes Mittel der SD-Kontrolle (EHLERS, 1973, p. 51).

3.4 Kontrolle durch Instruktion

Als wirksame Kontrolle sozialer Erwünschtheitstendenzen erscheint es vielen Autoren, diesem Response Set mittels Instruktion einen alternativen „Set“ entgegenzusetzen, z. B. in Gestalt der Aufforderung, ganz besonders ehrlich zu antworten und sich nicht davon leiten zu lassen, welche Antwort sozial erwünscht und welche weniger erwünscht sei. Jeder Experimentator weiß, daß Instruktionen kognitive Berge versetzen können, doch scheinen die Kräfte einer solchen vorwarnenden, immunisierenden Instruktion begrenzt zu sein, vielleicht weil sie in den meisten Instruktionen diagnostischer Verfahren bereits routinemäßig enthalten ist. Selbstverständlich bestehen Unterschiede in der Wirkung der einfachen Bitte, „ganz offen und ehrlich“ zu antworten und einer verschärften Zusatzinstruktion gegen Verfälschungstendenzen wie etwa derjenigen bei HOETH & KÖBLER (1967). Als effektiv zumindest für die Zuschreibung von Erwünschtheits-Indizes zu Statements oder Fragen haben sich allerdings *Faking*-Instruktionen,

d. h. Aufforderungen zur bewußten Verstellung erwiesen: Bittet man die Vpn, sich vorübergehend idealisiert, so günstig wie möglich einzuschätzen, so lassen sich auf diese Weise Items identifizieren, die besonders *faking*-anfällig und damit vermutlich besonders erwünschtheitsanfällig sind (vgl. z. B. WIGGINS, 1959; EYSENCK & EYSENCK, 1963; STOLLAK, 1965; HOETH, KUCKLICK & SIMMAT, 1965; GORMAN, 1968; LENNERTZ, 1969; EGGERT, 1971; VELICER & WEINER, 1975; FARLEY & GOH, 1976; HACKER et al., 1977; ANGLEITNER, 1976; ANGLEITNER et al., 1976). Mit dieser Methode ließ sich beispielsweise auch zeigen, daß diagnostische Verfahren, die mittels Forced-Choice-Technik SD-kontrolliert sein sollen, gegenüber *Faking*-Instruktionen nach wie vor anfällig sind (STOLLAK, 1965), daß Neurotizismus-Messungen verfälschbarer als Extraversions-Messungen sind (FARLEY & GOH, 1976; ANGLEITNER, 1976), usw. SCHNEIDER & HÜBNER (1980) verglichen in einem Experiment die Bearbeitungszeiten des FPI unter den *Faking*-Instruktionen (a) einen guten Eindruck zu machen, (b) einen schlechten Eindruck zu machen, mit (c) einer Kontrollgruppe unter Normalinstruktion; zur Herstellung eines „guten Eindrucks“ wurde die kürzeste, zur Erzeugung eines „schlechten Eindrucks“ dagegen die meiste Zeit aufgewendet. Daraus könnte man schließen, daß die üblichen Tempoinstruktionen („nicht lange nachdenken“, „zügig antworten“ etc.) sozialen Erwünschtheitstendenzen keineswegs abträglich, wie zumeist angenommen, sondern eher förderlich sein können.

Daß eine *Faking*-Instruktion des FPI vom Typus „Bewerbungssituation“ zu Antwortmustern führte, die einer einfachen „Ideal“-Instruktion sehr ähnlich sind, zeigte WITTICH (zit. n. FAHRENBERG, SELG & HAMPEL, 1978). Um die Wirkung von Instruktion und Untersuchungssituation zu studieren, ließen wir in einer unveröffentlichten Arbeit 20 Vpn (a) sich in eine Bewerbungssituation versetzen, (b) sich in einer Gruppensituation beim Urteilen so einstellen, daß man möglichst günstig eingeschätzt würde, (c) das gleiche in einer Laborsituation, isoliert von anderen Personen, tun. Zu beurteilen waren ethnische Gruppen anhand von Eigenschaftsbegrif-

fen; da die SDSV-Werte dieser Eigenschaften aus mehreren Untersuchungen aus der Literatur bekannt waren, wurde für jede abhängige Variable ein „Literatur-SD-Wert“ ermittelt und in die unter den Bedingungen (a) bis (c) angewendete Urteilsskala transformiert. Der Literatur-SD-Wert korrelierte nur mit den Werten unter den Ideal-Instruktionen (b) und (c) um 0.90; für die Bewerbungs-Bedingung (a) waren die Korrelationen je nach Urteilsobjekt erheblich niedriger. Varianzanalysen zeigten zusätzlich je nach ethnischer Gruppe, die beurteilt wurde, mehr oder weniger ausgeprägte Beurteilungsunterschiede unter den für die SD-Kontrolle durch Instruktionsmanipulation scheinbar gleichermaßen günstigen Bedingungen. SD-Kontrolle durch Instruktion scheint also zwar, worauf bereits EHLERS (1973) kurz hinwies, für bestimmte Zwecke erfolgversprechend, weil es die komplizierten, wenig ökonomischen und oft ineffektiven Prozesse der Item- und Antwortselektion umgeht und nur auf die allgemeine Versuchseinstellung (set) der Vpn abhebt. Verbesserungen und Verfeinerungen erscheinen aber angebracht, wenn sich nicht auch diese Kontrollmöglichkeit abnutzen soll.

4. Ein Verfahren der experimentellen Kontrolle sozialer Erwünschtheit durch Manipulation der Rahmensituation

Abseits von Persönlichkeitsforschung und Psychologischer Diagnostik, auf dem Gebiet der sozialpsychologischen Einstellungsmessung, wurden Überlegungen und Experimente zur Kontrolle sozialer Erwünschtheitstendenzen unternommen, die einen Spezialfall der Kontrolle durch Instruktionsmanipulation darstellen. Solche Versuche zielen ab auf die *Veränderung der gesamten Rahmensituation* der Einstellungs- bzw. Persönlichkeitsmessung, d. h. auf die Herstellung einer Untersuchungssituation, in der die Reduktion von Reaktionstendenzen durch Anstrengungen der Vp selbst minimiert werden. Die Vpn werden dabei unter den Laboratoriumsbedingungen scheinbarer physiologischer Messung in einen Zustand erhöhter Selbstaufmerksamkeit

versetzt, der es für sie vordringlich macht, bei Selbstberichten über eigenes Verhalten zutreffender, d. h. ehrlicher zu antworten und bei Reaktionen auf Einstellungsobjekte stärker individuellen Normen als allgemeinen Erwünschtheitsnormen zu gehorchen. Da die Methode mit *Täuschung* arbeitet, entstehen forschungsethische Probleme; auf diese soll im nächsten Abschnitt eingegangen werden. Denn das Bemühen des Versuchsleiters, Täuschungs- bzw. Verfälschungstendenzen seitens der Vp durch eigene Täuschungsstrategien entgegenzuwirken, wird bei dieser Methode der SD-Kontrolle besonders deutlich.

Der als *Bogus-Pipeline-Paradigma* bezeichnete methodische Ansatz wurde Anfang der 70er Jahre von E. E. JONES und seinen Mitarbeitern zur Reduzierung von SD-Effekten entwickelt (vgl. JONES & SIGALL, 1971).

Eine relativ aufwendige experimentelle Rahmensituation mit dem scheinbaren Ziel der Erforschung psychophysiologischer Beziehungen wird erstellt, in der die Vp durch falsches physiologisches Feedback motiviert werden soll, relativ SD-freie Antworten zu produzieren. Im ersten Teil der Versuchsprozedur lernt die Vp, daß ein eindrucksvoll aussehendes, als „adaptierter Elektromyograph“ (EMG) bezeichnetes Gerät tatsächlich und überzeugend die unwillkürlichen Muskelschwingungen in beiden Unterarmen registriert, während die Vp ein Steuerrad (in unseren eigenen Experimenten zwei Hebel bzw. Allrichtungstasten) festhält und sich dabei auf eine Beurteilungsskala von -3 (links) bis +3 (rechts) zu konzentrieren hat. Die Täuschung über die genannte „Fähigkeit“ des Apparates wird so vorgenommen, daß ein Gehilfe des Versuchsleiters Reaktionen der Vp auf einige unverfängliche, offensichtlich nicht besonders SD-anfällige Statements (über Film, Musik, Autos, Sport usw.), die zuvor im Papier-Bleistift-Verfahren erfaßt worden sind, kopiert und in der Validierungs- bzw. „Adaptations“-Phase des Elektromyographen, in der die gleichen Statements noch einmal am EMG beantwortet werden sollen, in den Computer eingibt, so daß die Vp anhand des Computer-Output lernt, daß der Apparat Reaktionen registriert, die annäherungsweise ihren eigenen entsprechen. Dieses Täuschungsverfahren, im Detail beschrieben bei SIGALL & PAGE (1971) bzw. MUMMENDEY, SCHIEBEL, TROSKE, HESENER & BOLTEN (1979), führte sowohl in den amerikanischen als auch in den eigenen Untersuchungen dazu, daß nahezu alle Vpn vorübergehend von der Leistungs- und Glaubwürdigkeit der physiologischen Registrierung überzeugt werden können.

Im zweiten, entscheidenden Abschnitt des Bogus-Pipeline-Experiments werden dann Einstellungsmessungen vorgenommen. (Gewöhnlich sollen Einstellungsobjekte in substantivischer Form mittels Adjektiven auf der Sieben-Punkten-Skala beurteilt werden; wir haben jedoch die Technik für die Beantwortung üblicher Fragebogenitems adaptiert.) Dabei wird eine zweite Täuschung eingeführt, indem die Vp instruiert wird, daß es in dem Versuch darum gehe, zu untersuchen, in welchem Maße jemand seine wirklichen Einstellungen und Bewertungen kenne bzw. in welchem Maße jemand sein eigenes Verhalten richtig wahrnehme. Die Vp soll demgemäß die Werte, die auf dem EMG-Output erscheinen, bei abgedecktem Monitor schätzen bzw. vorhersagen. Es wird in Aussicht gestellt, daß später Übereinstimmungskoeffizienten zwischen dem subjektiven Urteil und der scheinbar objektiven Registrierung berechnet werden. Die Schätzungen der Vp über ihre eigenen, scheinbar objektiv gemessenen Reaktionen werden dann als relativ SD-verzerrungsfreie Messungen betrachtet, ohne daß – was ja auch gar nicht möglich wäre – ein Übereinstimmungsmaß berechnet wird. Im Anschluß an die Bogus-Pipeline-Messung werden die Vpn über den Versuch aufgeklärt.

Erste Hinweise auf die Überlegenheit des Bogus-Pipeline-Paradigmas über herkömmliche Papier-Bleistift-Messungen in bezug auf die Verminderung sozialer Erwünschtheitsreaktionen brachten SIGALL & PAGE (1971) mit einer Untersuchung von Urteilen über Farbige und weiße Amerikaner. Der mittlere Grad an sozialer Erwünschtheit der den beiden Volksgruppen zugeordneten Urteile war unter Papier-Bleistift-Bedingungen fast gleich (0.53 für Amerikaner, 0.49 für Neger), während unter Bogus-Pipeline-Bedingungen die „Amerikaner“ einen mittleren SDSV-Wert von 0.84, die Farbigen einen Wert von -0.03 erhielten. Nur in der durch scheinbare physiologische Messung bestimmten experimentellen Situation wurde also gewissermaßen zugegeben, daß man Farbige ungünstiger einschätzt als die Angehörigen der eigenen Gruppe. Eine Reihe von Replikationen und weiteren Untersuchungen von Urteilen über Gegenstände, die dazu geeignet sind, SD-Reaktionen zu fördern (z. B. Beurteilungen im Rahmen des Ähnlichkeits-Attraktions-Paradigma; vgl. BYRNE, 1971) hat im Laufe der Zeit die Vermutung bestätigt, daß die Herstellung einer verschärften Urteils-Rahmensituation im Bogus-Pipeline-Paradigma zur Verringerung von SD-Tendenzen

geeignet ist (vgl. zusammenfassend BRACKWEDDE, 1980). Über die im wesentlichen zufriedenstellenden Ergebnisse einer deutschen Replikation, bei der vor allem als Einstellungsobjekte Türken und Westeuropäer verglichen wurden, wurde auf der 22. Tagung der experimentell arbeitenden Psychologen in Tübingen berichtet (MUMMENDEY, 1980).

Für die Annahme, in der Bogus-Pipeline-Situation komme es ganz allgemein zu einer Veränderung der Anpassung an soziale Normen, liegt bislang keine elaborierte theoretische Begründung vor. Geht man einmal davon aus, daß durch die in der speziellen Situation gegebene Instruktion an die Vp, sie solle so antworten, daß dies mit ihren „wirklichen Einstellungen“ übereinstimmt, sie solle sich bemühen, „ihre eigenen Gefühle zu erkennen“, „ihr eigenes Verhalten richtig wahrzunehmen“, usw., eine intensive *Lenkung der Aufmerksamkeit auf die eigene Person* erfolgt, so kann man den intendierten – und vermutlich auch erzielten – Zustand der Vp als einen solchen der selbstzentrierten Aufmerksamkeit oder – in der Terminologie von DUVAL & WICKLUND (1972) – der objektiven Selbstaufmerksamkeit bezeichnen. Insofern wäre das Anschließen der Vp an den scheinbaren Elektromyographen funktional äquivalent dem Aufstellen eines Spiegels, der Ankündigung fotografiert zu werden, der Wiedergabe der eigenen Stimme vom Tonband oder der Konfrontation mit einem Publikum, also eben jenen mehr oder weniger wohlfeilen Hilfsmitteln, deren sich die Befürworter der Theorie der objektiven Selbstaufmerksamkeit in Experimenten bedienen, um eine Person besonders selbstaufmerksam zu machen (vgl. WICKLUND, 1975, 1979a, 1980).

Da in mehreren Untersuchungen gezeigt werden konnte, daß Personen im Zustande erhöhter objektiver Selbstaufmerksamkeit eher in Übereinstimmung mit ihren zuvor gemessenen Einstellungen urteilen, so daß es bei ihnen auch zu besserer Einstellungs-Verhaltens-Konsistenz kommt, wäre das Konzept der objektiven Selbstaufmerksamkeit ein geeigneter Ansatz zur Erklärung von SD-Kontrolle im Bogus-Pipeline-Experiment. Allerdings erhöht sich die Konsi-

stanz von Verhalten und Urteilen einer Person anscheinend nur in dem Maße, in dem diese Urteile mit sozialen Normen übereinstimmen. Als Beispiele hierfür mögen die Versuche zum prosozialen Verhalten dienen, über die WICKLUND auf dem 22. Internationalen Kongreß für Psychologie in Leipzig berichtete; sie zeigen für amerikanische und osteuropäische Verhältnisse in konsistenter Weise, daß altruistische Akte im Zustand erhöhter Selbstaufmerksamkeit nur dann häufiger auftreten, wenn die soziale Norm, d. h. anderen zu helfen, hochgradig *salient* ist. Entscheidend für die Interpretierbarkeit von Bogus-Pipeline-Effekten im Sinne des Konzepts der objektiven Selbstaufmerksamkeit wird damit, ob die Vp in der speziellen experimentellen Situation eher in Richtung auf individuelle oder in Richtung auf Gruppen-Normen urteilt; im letzteren Falle wären es ja gerade die unerwünschten SD-Tendenzen, die im Bogus-Pipeline-Paradigma unterstützt würden. Von WICKLUND (1980, p. 203 f.) referierte Untersuchungsergebnisse weisen teils auf die Zunahme, teils auf die Abnahme von Konformität bei erhöhter objektiver Selbstaufmerksamkeit hin.

Führt man als weiteres Konzept dasjenige des sozialen Drucks ein und postuliert, daß das Individuum zur Vermeidung von Inkonsistenz im Zustand erhöhter Selbstaufmerksamkeit dem stärksten in der Situation aktuellen Druck nachgibt, so läßt sich unter Berücksichtigung der spezifischen Charakteristika der Bogus-Pipeline-Rahmensituation gut begründen, warum hier interne, individuelle Standards stärker oder salienter als Gruppenstandards werden: Die Person wird vom Versuchsleiter nachdrücklich auf ihre „wirklichen eigenen“ Wahrnehmungen und Urteile hingewiesen, es ist – außer der Maschine – nur der Versuchsleiter vorhanden, der betont, daß man sich zu wissenschaftlichen Zwecken allein für individuelle Besonderheiten interessiere, usw. Auf diese Weise ließe sich erklären, wieso in der durch das Bogus-Pipeline-Paradigma geschaffenen Rahmensituation für individuelle Urteile Antworten wahrscheinlich werden, die stärker individuellen als allgemeinen Erwünschtheits-Normen gehorchen. Experimentelle Belege für die Angemessenheit eben dieser

theoretischen Erklärung stehen allerdings noch aus. Inzwischen haben wir lediglich versucht, in einer direkten, auch von den amerikanischen Vorbildern nicht geleisteten Weise die Frage zu entscheiden, ob in der Bogus-Pipeline-Situation tatsächlich „Soziale Erwünschtheit“ im herkömmlichen, auf die Unterscheidung von Fragebogen-Inhalt und -Stil (vgl. JACKSON & MESSICK, 1958) bezogenen Sinne reduziert wird. Dazu wurden die Antworten unabhängiger Stichproben studentischer Vpn auf die Items mehrerer deutschsprachiger SD-Fragebogen sowohl unter üblichen Papier-Bleistift-Bedingungen als auch unter Bogus-Pipeline-Bedingungen registriert und miteinander verglichen (MUMMENDEY & BOLTEN, 1981). Es konnte gezeigt werden, daß in der Bogus-Pipeline-Situation signifikant weniger SD-Antworten gegeben werden als bei der üblichen Fragebogen-Applikation.

Die Konstruktion einer experimentellen Rahmensituation, in der die Vp lernt, sich bei der Beantwortung von Statements oder Fragen, beim Abgeben von Urteilen unterschiedlichster Art an relativ stärker individuell als gruppenspezifisch bestimmten Normen zu orientieren, erscheint für die Zukunft nur aussichtsreich, wenn ihre Praktikabilität erhöht und ihr zugleich der Charakter des Spektakulären genommen werden kann. Dabei handelt es sich zum Teil um Probleme, die bereits in forschungsethische Fragen der Anwendung und Verbreitung von Täuschungsverfahren hineinreichen.

5. Ethische Probleme der Kontrolle sozialer Erwünschtheit

Kurioserweise, so könnte man sagen, scheint dasjenige, was *sozial* erwünscht ist, unter Gesichtspunkten *psychologischer* Forschung unerwünscht zu sein. Es herrscht offensichtlich eine gewisse Zweideutigkeit in der Bewertung des Verhaltens von Versuchspersonen bzw. Teilnehmern an psychologischen Untersuchungen, wenn man soziale Normen, die wir alle normalerweise akzeptieren, mit psychologischen For-

schungsnormen konfrontiert. Die Kuriosität läßt sich natürlich damit erklären, daß man postuliert, es sei grundsätzlich zwischen wahrem Inhalt und Reaktionstendenz, also zwischen Substanz und Erscheinung, zu trennen, und der Forscher müsse vorübergehend tricksen, um an die Substanz eines subjektiven Urteils zu gelangen. Da er dabei jedoch – insbesondere deutlich im Bogus-Pipeline-Paradigma – mit Täuschung (deception) arbeitet, muß er sich die Fragen nach der Ethik gefallen lassen. Diese Fragen werden konventionellerweise von interessierten Psychologiekritikern aus dem Umkreis der Sozialwissenschaften gestellt, sie werden jedoch neuerdings – und auch hierzulande – von einer Reihe experimentell und empirisch arbeitender Psychologen selbst häufiger aufgeworfen (vgl. z. B. KELMAN, 1967; SCHLENKER & FORSYTH, 1977; SCHWITZGEBEL & KOLB, 1978; STOLZ & Associates, 1978; WEST & GUNN, 1978; DIENER & CRANDALL, 1978; AMERICAN PSYCHOLOGICAL ASSOCIATION, 1979; IRLE, 1979; SCHULER, 1980; KUMPF, im Druck).

Gegenstand unserer kurzen Erörterung sollen hier nicht die öffentlichkeitswirksamen Experimente, etwa zum Gehorsamsverhalten (MILGRAM, 1963) oder zur Verhaltensbeobachtung in Toiletten (MIDDLEMEST, KNOWLES & MATTER, 1976) sein, die an anderer Stelle unter ethischen Aspekten diskutiert worden sind (BAUMRIND, 1964; MILGRAM, 1964; KOOSCHER, 1977), sondern lediglich die Deception- bzw. Täuschungs-Problematik, wie sie im Prinzip bei allen Versuchen, soziale Erwünschtheit zu kontrollieren, auftritt.

Täuschungen der Vp werden vor allem bei sozialpsychologischen Experimenten, aber auch bei vielen diagnostischen Routineverfahren vorgenommen, wenn die bewußte Kenntnis der kritischen Variablen einer Untersuchung das Verhalten der Vp wesentlich beeinflussen könnte. Täuschung kann als *Hilfsmittel zur Abwendung von Vpn-Effekten* (vgl. ORNE, 1962; ROSENBERG, 1965) angesehen werden, die sich durch das Bemühen der meisten Vpn, eine „gute Vp“ zu sein und die Anforderungscharakteristika (demand characteristics) eines Versuchs zu erkennen, einstellen. Ihr unbedachter und gleichsam

sportlicher Einsatz wird von fast allen einschlägigen Autoren abgelehnt, und es wird gewöhnlich dafür plädiert, mit dem Mittel der Täuschung sparsam umzugehen (vgl. IRLE, 1979). Auf der anderen Seite scheinen bestimmte Typen von Experimenten und Untersuchungen – beispielsweise zur kognitiven Dissonanz oder zur Attribution von Erregung – mit der Anwendung von Täuschung zu stehen und zu fallen; die gelegentlich als Alternativen empfohlenen Methoden des Rollenspiels, der Feldforschung usw. vermögen doch nur sehr selten die Anwendung experimenteller Versuchspläne oder die Verwendung von Tests mit manipuliertem Grad an Durchschaubarkeit zu ersetzen.

Einige *Gegenargumente* zum Gebrauch von Täuschung können lauten:

- Unter Ausnutzung seiner sozialen Macht erschleicht sich der Forscher auf betrügerische Weise wissenschaftliche Erkenntnisse (vgl. KELMAN, 1972)
- Durch falsche Rückmeldung an die Vp wird diese unter Umständen nachhaltig geschädigt (vgl. KUMPF, im Druck)
- Täuschung im psychologischen Experiment kann negative Auswirkungen auf Vpn im Sinne der Modellierung von Unaufrichtigkeit haben (BAUMRIND, 1964)
- Bei der getäuschten Vp kann Mißtrauen entstehen, das für weitere Experimente und die gesamte psychologische Forschung schädlich sein kann (vgl. KUMPF, im Druck)
- Das Wissen um die Täuschung hat sich in einigen Untersuchungen in geringerem Maße als erwartet ausgewirkt (vgl. SCHULER, 1980).

Als Argumente *für* die Anwendung von Täuschung können u. a. dienen:

- Täuschung ist aus methodologischen Gründen, z. B. zur Erhaltung von Nonreaktivität, Spontaneität etc. der Vp. geboten (vgl. MÉTRAUX, im Druck)
- Täuschung ist in der psychologischen Forschung üblich (vgl. SCHULER, 1980)
- Täuschung geschieht nur vorübergehend, da gewöhnlich postexperimentell über die Art der Untersuchung aufgeklärt wird
- Viele Vpn beurteilen Täuschung nachträglich als keineswegs ethisch anfechtbar; sie beweren

ten Täuschung oft weniger negativ als Psychologen (vgl. KUMPF, im Druck)

- Täuschung könnte Teil einer impliziten (oder sogar expliziten) Vereinbarung zwischen Vp und VI sein, wenn die Vp sich auf ein psychologisches Experiment oder einen Test einläßt (vgl. SCHWITZGEBEL & KOLB, 1978)
- Vollständige Offenheit und Ehrlichkeit ist eigentlich nirgendwo im öffentlichen Leben üblich (vgl. SCHULER, im Druck)
- Es ist bislang nichts über tatsächlich eingetretene Schädigungen durch die Anwendung von Täuschung bekannt (vgl. KUMPF, im Druck).

Angeichts der begründet erscheinenden relativen Notwendigkeit der Anwendung von Täuschung könnte die Einhaltung der einschlägigen Richtlinie aus dem Verhaltenskatalog von DIENER & CRANDALL (1978) in Verbindung mit den APA-Richtlinien in der Version von 1979 (AMERICAN PSYCHOLOGICAL ASSOCIATION, 1979), und zwar den Subprinzipien (d) und (h) angezeigt sein: Zunächst sollte Täuschung nur nach einer ethischen Analyse der Forschungssituation praktiziert werden. Wenn dann die methodischen Erfordernisse Täuschung als notwendig erscheinen lassen, soll der Forscher die Versuchsteilnehmer so bald wie möglich aufklären und ihnen zusätzlich die Gründe für die Vorgehensweise und ihre Rechtfertigung erläutern. Es sollte demnach nicht genügen, das allgemein anerkannte Prinzip der Freiwilligkeit der Teilnahme von Vpn als ausreichend zu erachten, da Freiwilligkeit und selbst Teilnahme gegen Bezahlung nicht implizieren, daß die Vp sich verpflichtet, jede Art von Untersuchung bis zu ihrem Ende mitzumachen.

Die Idee einer *gegenseitigen Verpflichtung von VI und Vp* kann jedoch auch folgendes implizieren: Die Vp weiß oder kann sich denken, daß der VI zu wissenschaftlichen Zwecken an ihren „tatsächlichen“ bzw. „aufrichtigen“ Reaktionen interessiert ist. Stellt es nicht insofern eine Art Bruch des Kontraktes zwischen Vp und VI dar, wenn die Vp beispielsweise eher in sozial erwünschter, beschönigender Weise antwortet? So wehrt sich z. B. KELMAN (1968) dagegen, aus ähnlichen Erwägungen das Täuschungsproblem

einseitig zu sehen. Täuschung könnte demgemäß ein – wenn man so will – Mittel der Prophylaxe gegen mögliche Brüche der Übereinkunft zwischen beiden Partnern durch sozial erwünschte Reaktionen sein; indem ein solcher Bruch der Vereinbarung als menschlich, allzumenschlich antizipiert wird, wird versucht, ihm durch einen andersgearteten Bruch, nämlich die Anwendung von Täuschung, entgegenzuwirken. Schließlich verdient der Hinweis SCHULER's (1980, p. 51) auf „glaubhafte und mutmaßlich unschädliche Erklärungen“ bei der Täuschung Beachtung: Wenn schon mit Täuschung gearbeitet wird, so sollten die falschen Erklärungen sowohl glaubwürdig als auch unschädlich sein.

Da der ausgezeichneten Abhandlung des Ethik-Problems durch SCHULER (1980) nur wenig hinzuzufügen ist, da die APA-Normen sich seit den ersten Beratungen eines Ethik-Ausschusses im Jahre 1938 nicht gewandelt haben (ebenso unverändert blieben die berufsethischen Verpflichtungen des Berufsverbandes Deutscher Psychologen, die auf das Täuschungsproblem explizit gar nicht eingehen), bleibt vorerst nur folgende Feststellung:

Ethisch bedenklich (z. B. im Sinne des Prinzips Nr. 8a der APA-Richtlinien) wäre es sicherlich, wenn zur Kontrolle sozialer Erwünschtheitstendenzen Vpn im Bogus-Pipeline-Verfahren mit einem echten und funktionierenden EMG-Gerät, das dann einer Art Lügendetektor entspräche, untersucht würden und dies den Vpn verschwiegen würde. Ein Kontrollverfahren wie das Bogus-Pipeline-Experiment verfährt aber in genau umgekehrter Weise: Mit einem nicht funktionierenden Gerät wird eine Kontrolle physiologischer, nicht steuerbarer Reaktionen vorgetäuscht. Überträgt man diesen Gedanken auf die Papier-Bleistift-Verfahren zur Kontrolle sozialer Erwünschtheit, so könnte man sogar sagen: Ethisch bedenklich wäre es vielleicht, beim Versuch der Kontrolle von Reaktionstendenzen in Persönlichkeitsforschung und Psychologischer Diagnostik Itemformen, Antwortkombinationen oder spezielle Kontrollskalen einzusetzen, die tatsächlich funktionieren bzw. vollständig effizient sind. Zum gegenwärtigen Zeitpunkt stellen diese Verfahren jedoch Hilfsmittel dar,

die nur teilweise wirksam sind, so daß einige Autoren vermutlich zu Recht von ihrer Anwendung abgekommen sind und die restlichen Autoren einstweilen von allzu starkem Gewissensdruck entlastbar erscheinen.

6. Die Anerkennung sozialer Erwünschtheit in Teilbereichen der psychologischen Forschung

Abschließend soll die Frage wenigstens aufgeworfen werden, ob die Psychologie, insbesondere die Differentielle und Diagnostische Psychologie ohne das Konzept sozial erwünschter Reaktionen auskommen kann. Die Frage muß präzisiert werden, da eingangs gezeigt wurde, daß die Sensibilität einer Person für den Erwünschtheitsgrad eines Items als fundamentale Bestimmungsgröße des Reagierens auf verbale Stimuli überhaupt angesehen werden kann. Die Frage stellt sich vielmehr so: ob Forschungsfragestellungen denkbar und realisierbar sind, die die Auffassung von Social-Desirability-Varianz als Fehlervarianz unnötig machen und so gleichzeitig dazu führen, daß sich Täuschungsstrategien seitens der untersuchten Person wie auch des Untersuchers erübrigen.

Einen Zugang zu Persönlichkeitsunterschieden im Sinne einer willkommenen, ja geradezu notwendigen Berücksichtigung sozialer Erwünschtheitstendenzen wird man gewinnen wollen, (a) wenn kein Wahrheitskriterium vorhanden ist, d. h. wenn sich nicht entscheiden läßt, ob eine Antwort mehr oder weniger zutreffend ist – also niemals beim Bericht über manifestes Verhalten, sondern stets allenfalls bei Berichten über Einstellungen bzw. Wahrnehmungen und Bewertungen, (b) wenn die Vermischung der individuellen Stellungnahme mit einer sozialen Norm mehr oder weniger Gegenstand des Forschungsinteresses ist, d. h. wenn ermittelt werden soll, wie das Individuum einen Gegenstand unter Einschluß der Perspektive seiner Bezugsgruppe perzipiert und evaluiert.

Vermutlich treffen diese beiden Voraussetzungen gegenwärtig weitgehend erst auf Teile der *Selbstkonzept*-Forschung zu. Werden beispielsweise Persönlichkeitsfragebogen nicht als objektive Tests im CATTELLSchen Sinne, sondern als Beschreibungsverfahren von Selbstkonzepten von Personen aufgefaßt, so liegen nicht etwa Merkmale des Individuums im Forschungsinteresse, die intersubjektiv-konkordant erfaßbar, also objektivierbar und eigentlich nur aus forschungsökonomischen Gründen über subjektive Befragungsverfahren erhebbar sind. Vielmehr produzieren Individuen dann, wenn sie über ihr ihnen als real erscheinendes Selbstbild Auskunft geben sollen (selbstverständlich erst recht im Fall der Erfassung des idealen Selbstbildes), prinzipiell nicht nachprüfbar Urteile über die eigene Person, die vermutlich weitgehende Anteile dessen enthalten, was für das Individuum relevante soziale Gruppen bejahen würden. Bei derartigen Selbstbeurteilungen bzw. Selbsteinstellungen interessieren den Forscher durchaus die internalisierten – oder auch nur als Reaktion auf spezifische Anforderungsmerkmale psychologischer Forschungssituationen aktualisierte – Orientierungen an Gruppennormen als selbstverständlicher Bestandteil von Selbstbildern; sie sind von Response-Set-Kontrollen ausgenommen, da es für sie keine Außenkriterien gibt. Es interessiert dabei weniger, ob oder in welchem Maße die Versuchsperson sich selbst oder dem Versuchsleiter etwas vormacht, sondern es interessiert das von der Vp produzierte Selbstbild, „wie es sich von sich her zeigt“ (HUSSERL). Da eine verhaltensbeeinflussende Funktion von Selbstkonzepten allgemein akzeptiert zu sein scheint (vgl. z. B. WICKLUND, 1979 b; vgl. weitere Beiträge in FILIPP, 1979; vgl. zur Auffassung des Selbstbildes als übergeordnete Verhaltensnorm auch SCHULER, 1980, p. 48 f.), kann allgemein von gewissen Prädiktoreigenschaften von Selbstkonzept-Maßen für menschliches Verhalten auch dann ausgegangen werden, wenn die gemessenen Aspekte oder Ausprägungsgrade des Selbstkonzepts beispielsweise erhebliche Anteile an Wunschenken enthalten, also mehr oder weniger „realistisch“ bzw. unrealistisch sind.

Ein verstärkter Ausbau der Selbstkonzeptfor-

schung wie überhaupt eine Überbetonung kognitiver Prozesse auf seiten der Versuchspersonen bis hin zu einer unkritischen Übernahme der Vp-Perspektive als einzig gültiger psychologischer Erkenntnisquelle wird jedoch nicht empfohlen, da dies letztlich zu einer Renaissance introspektiver Psychologie mit allen längst bekannten Nachteilen führen könnte. Zumindest in Teilbereichen von Differentieller und Diagnostischer Psychologie könnte jedoch möglicherweise das aufgezeigte Wechselspiel von Täuschung (durch die Vp) und Enttäuschung (durch den VI) durchbrochen werden, indem die Bedeutung der Reaktionstendenz „soziale Erwünschtheit“ dadurch verringert wird, daß man ihre Faktizität – und zugleich ihre psychologische Notwendigkeit, z. B. für Selbstkonzepte – anerkennt. In denjenigen Teilbereichen bzw. Forschungsfeldern jedoch, auf denen das Konzept sozial erwünschter Reaktionen auf klassische Art und Weise als psychologisch unerwünschter Fehler bzw. Bias beibehalten werden muß, sollte dies in der differenziertest-möglichen Form geschehen.

Literatur

- AHAMMER, I. M.: Desirability judgments as a function of item content, instructional set, and sex: A life-span developmental study. *Human Development* 1971, 14, 195 – 207.
- AHAMMER, I. M., & Baltes, P. B. Objective versus perceived age differences in personality: How do adolescents, adults, and older people view themselves and each other? *Journal of Gerontology* 1972, 27, 46 – 51.
- ALLAMAN, J. D., JOYCE, C. S., & CRANDALL, V. C.: The antecedents of social desirability response tendencies of children and young adults. *Child Development* 1972, 43, 1135 – 1160.
- AMERICAN PSYCHOLOGICAL ASSOCIATION.: *Ethical Standards of Psychologists (1979 Revision)*. Washington, D. C.: American Psychological Association, 1979.
- ANDERSON, N. H.: Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology* 1968, 9, 272 – 279.
- ANGLEITNER, A.: *Methodische und theoretische Probleme bei Persönlichkeitsfragebogen unter besonderer Berücksichtigung neuerer deutschsprachiger Fragebogen*. Unveröffentlichte Habilitationsschrift, Universität Bonn, 1976.
- ANGLEITNER, A., STUMPF, H. & WIECK, T.: Die „Personality Research Form“ von Jackson: Konstruktion, bisheriger Forschungsstand und vorläufige Ergebnisse zur Äquivalenzprüfung einer deutschen Übersetzung. *Wehrpsychologische Untersuchungen* 1976, 3.
- BAUMRIND, D.: Some thoughts on ethics of research: After reading Milgram's "Behavioral Study of Obedience". *American Psychologist* 1964, 19, 421 – 423.
- BENTLER, P. M., JACKSON, D. N., & MESSICK, S.: Identification of content and style: A two-dimensional interpretation of acquiescence. *Psychological Bulletin* 1971, 76, 186 – 204.
- BLOCK, J.: *The challenge of response sets*. New York: Appleton-Century-Crofts, 1965.
- BÖCHER, W.: Motivations- und situationsspezifische Einflüsse auf die Ergebnisse von Fragebogenuntersuchungen. In: Merz, F. (Hrsg.): *Bericht über den 25. Kongreß der Deutschen Gesellschaft für Psychologie in Münster 1966*. Göttingen: Hogrefe, 1967.
- BRACKWEDE, D.: Das Bogus-Pipeline-Paradigma: Eine Übersicht über bisherige experimentelle Ergebnisse. *Zeitschrift für Sozialpsychologie* 1980, 11, 50 – 59.
- BROWN, P. K.: The social desirability variable and verbal learning performance. *Journal of Educational Psychology* 1960, 51, 52 – 59.
- BUXBAUM, O.: *Die soziale Erwünschtheit von Persönlichkeitseigenschaften*. Berichte aus dem Institut für Psychologie der Universität Graz, Universität Graz, Juni 1976.
- BUXBAUM, O.: *Die soziale Erwünschtheit von Persönlichkeitseigenschaften (1979)*. Unveröffentlichtes Manuskript, Universität Graz, Institut für Psychologie, 1979.
- BYRNE, D.: *The attraction paradigm*. New York: Academic Press, 1971.
- COWEN, E. L., BUDIN, W. & BUDIN, F. A.: The social desirability of trait descriptive terms: A paired-comparison approach. *Journal of Social Psychology* 1964, 63, 265 – 279.
- CRANDALL, V. C., CRANDALL, V. C. & KATKOVSKY, W.: A children's social desirability questionnaire. *Journal of Consulting Psychology* 1965, 29, 27 – 36.
- CRONBACH, L. J.: Response sets and test validity. *Educational and Psychological Measurement* 1946, 6, 475 – 494.
- CRONBACH, L. J.: Further evidence on response sets and test design. *Educational and Psychological Measurement* 1950, 10, 3 – 31.
- CROTT, H. W., PRÜFER, P. & WOLFSHÖRNDL, H.: Erwünschtheit von Persönlichkeitseigenschaften für verschiedene Altersgruppen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 1977, 9, 270 – 276.

- CROTT, H. W. & ROSSRUCKER, K.: Erwünschtheit von Eigenschaften in Abhängigkeit von Alter, Geschlecht und Schichtzugehörigkeit. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 1974, 6, 241–261.
- CROWNE, D. P. & MARLOWE, D.: A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology* 1960, 24, 349–354.
- CROWNE, D. P. & MARLOWE, D.: *The approval motive: Studies in evaluative dependence*. New York: Wiley, 1964.
- DICKEN, C.: Good impression, social desirability, and acquiescence as suppressor variables. *Educational and Psychological Measurement* 1963, 23, 699–720.
- DICKENBERGER, D., HOLTZ, S. & GNECH, G.: Bedürfnis nach sozialer Anerkennung: Validierung der „Marlowe-Crowne Social Desirability Scale“ über ein Konzept individuell relevanter Gruppen. *Diagnostica* 1978, 24, 24–38.
- DIENER, E. & CRANDALL, R.: *Ethics in social and behavioral research*. Chicago: The University of Chicago Press, 1978.
- DUVAL, S. & WICKLUND, R. A.: *A theory of objective self-awareness*. New York: Academic Press, 1972.
- EDWARDS, A. L.: The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology* 1953, 37, 90–93.
- EDWARDS, A. L.: *Edwards Personal Preference Schedule. Manual*. New York: Psychological Corporation, 1954.
- EDWARDS, A. L.: *The social desirability variable in personality assessment and research*. New York: Dryden Press, 1957. (a)
- EDWARDS, A. L.: *Techniques of attitude scale construction*. New York: Appleton, 1957. (b)
- EDWARDS, A. L.: *The measurement of personality traits by scales and inventories*. New York: Holt, Rinehart & Winston, 1970.
- EDWARDS, A. L. & ABBOTT, R. D.: Measurement of personality traits: Theory and technique. *Annual Review of Psychology* 1973, 24, 241–278.
- EGGERT, D.: Untersuchungen zur psychometrischen Eignung eines neuen Fragebogens der neurotischen Tendenz und der Extraversion von Eysenck (EPI). In: Duhm, E. (Hrsg.): *Praxis der klinischen Psychologie*, Band 2, Göttingen: Hogrefe, 1971.
- EHLERS, T.: Zur Effektivität der Kontrollen von Reaktionseinstellungen. In: Reinert, G. (Hrsg.): *Bericht über den 27. Kongreß der Deutschen Gesellschaft für Psychologie in Kiel 1970*. Göttingen: Hogrefe, 1973.
- ENDLER, N. S. & MAGNUSSON, D. (Eds.): *Interactional psychology and personality*. New York: Wiley, 1976.
- EPSTEIN, R.: Need for approval and the conditioning of verbal hostility in asthmatic children. *Journal of Abnormal and Social Psychology* 1964, 69, 105–109.
- EYSENCK, S. B. G. & EYSENCK, H. J.: An experimental investigation of „desirability“ response set in a personality questionnaire. *Life Sciences* 1963, 2, 343–355.
- FAHRENBERG, J., SELG, H. & HAMPEL, R.: *Das Freiburger Persönlichkeitsinventar FPI*. Göttingen: Hogrefe, 1978³.
- FARLEY, F. H. & GOH, D. S.: PENmanship: Faking the P-E-N. *British Journal of Social and Clinical Psychology* 1976, 15, 139–148.
- FILIPP, S.-H. (Hrsg.): *Selbstkonzept-Forschung. Probleme, Befunde, Perspektiven*. Stuttgart: Klett-Cotta, 1979.
- FISKE, D. W. & PEARSON, P. H.: Theory and technique of personality measurement. *Annual Review of Psychology* 1970, 21, 49–87.
- FORD, L., JR. & RUBIN, B. M.: A social desirability questionnaire for young children. *Journal of Consulting and Clinical Psychology* 1970, 35, 195–204.
- GOLDBERG, L. R., RORER, L. G. & GREENE, M. M.: The usefulness of „stylistic“ scales as potential suppressor or moderator variables in predictions from the CPI. Oregon Research Institute Research Bulletin, 10 (3), 1970 (zit. n. Wiggins, 1970).
- GORMAN, B. S.: Social desirability factors and the Eysenck Personality Inventory. *Journal of Psychology* 1968, 69, 75–83.
- HÄCKER, H., SCHWENKMEZGER, P. & UTZ, H.: Über die Verfälschbarkeit von Persönlichkeitstests unter SD-Instruktion und in einer Auslesesituation. In: Tack, W. H. (Hrsg.): *Bericht über den 30. Kongreß der Deutschen Gesellschaft für Psychologie in Regensburg 1976*, Band 2. Göttingen: Hogrefe, 1977.
- HERRMANN, T.: Die Eigenschaftskonzeption als Heterostereotyp. Kritik eines persönlichkeitspsychologischen Geschichtsklischees. *Zeitschrift für Differentielle und Diagnostische Psychologie* 1980, 1, 7–16.
- HOETH, F., BÜTTEL, R. & FEYERABEND, H.: Experimentelle Untersuchungen zur Validität von Persönlichkeitsfragebogen. *Psychologische Rundschau* 1967, 18, 169–184.
- HOETH, F. & GREGOR, H.: Guter Eindruck und Persönlichkeitsfragebogen. *Psychologische Forschung* 1964, 28, 64–88.
- HOETH, F. & KÖBLER, V.: Zusatzinstruktionen gegen Verfälschungstendenzen bei der Beantwortung von Persönlichkeitsfragebogen. *Diagnostica* 1967, 13, 117–130.
- HOETH, F., KUCKLICK, G. & SIMMAT, W.: Experimentelle Untersuchungen zum Problem des „guten Eindrucks“. *Zeitschrift für Experimentelle und Angewandte Psychologie* 1965, 12, 59–85.

- HORST, P.: *The prediction of personal adjustment*. New York: Social Science Research Council, 1941.
- IRLE, M.: Das Instrument der „Täuschung“ in der Verhaltens- und Sozialwissenschaftlichen Forschung. *Zeitschrift für Sozialpsychologie* 1979, 10, 305 – 330.
- JACKSON, D. N.: Balanced scales, item overlap, and the stables of augeas. *Educational and Psychological Measurement* 1967, 27, 502 – 507.
- JACKSON, D. N.: *Personality research form manual* (rev. ed.). Port Huron, Mi.: Research Psychologists Press, 1974.
- JACKSON, D. N. & MESSICK, S.: Content and style in personality assessment. *Psychological Bulletin* 1958, 55, 243 – 252.
- JANKE, W.: Das Dilemma von Persönlichkeitsfragebogen. Einleitung des Symposiums über Konstruktion von Fragebogen. In: Reinert, G. (Hrsg.): *Bericht über den 27. Kongreß der Deutschen Gesellschaft für Psychologie in Kiel 1970*. Göttingen: Hogrefe, 1973.
- JONES, E. E. & SIGALL, H.: The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin* 1971, 76, 349 – 364.
- KATZ, I., HENCHY, T. & ALLEN, H.: Effects of race of tester, approval-disapproval, and need on Negro children's learning. *Journal of Personality and Social Psychology* 1968, 8, 38 – 42.
- KEIL, W.: *Inhalt, Reaktionseinstellung und Stil. Meßtheoretische Untersuchungen zum Fragebogenverfahren*. Unveröffentlichte Dissertation, Universität Mainz, 1968.
- KEIL, W.: Reaktionseinstellung und Fragebogenkonstruktion. In: Reinert, G. (Hrsg.): *Bericht über den 27. Kongreß der Deutschen Gesellschaft für Psychologie in Kiel 1970*. Göttingen: Hogrefe, 1973.
- KELMAN, H.: Human use of human subjects: The problem of deception in social psychological experiments. *Psychological Bulletin* 1967, 1, 1 – 11.
- KELMAN, H. C.: *A time to speak: On human values and social research*. San Francisco: Jossey-Bass Inc., 1968.
- KELMAN, H. C.: The rights of the subject in social research: An analysis in terms of relative power and legitimacy. *American Psychologist* 1972, 27, 989 – 1016.
- KIRBY, D. & GARDNER, R.: *Norms of 208 words typically used in the assessment of ethnic stereotypes*. University of Western Ontario Research Bulletin, No. 177, 1971 (zit. n. Edwards & Abbott, 1973).
- KLAPPROTT, J.: Erwünschtheit und Bedeutung von 338 alltagspsychologischen Eigenschaftsbegriffen. *Psychologische Beiträge* 1972, 14, 496 – 520.
- KLEIN, P.: Soziale Erwünschtheit von Eigenschaften in Abhängigkeit von Nationalität, Schulbildung und Geschlecht der Beurteiler. *Psychologie und Praxis* 1974, 18, 86 – 92.
- KLETT, C. J.: The social desirability stereotype in a hospital population. *Journal of Consulting Psychology* 1957, 21, 419 – 421.
- KLETT, C. J. & YAUKEY, D. W.: A cross-cultural comparison of judgments of social desirability. *The Journal of Social Psychology* 1959, 49, 19 – 26.
- KOOCHER, G. P.: Bathroom behavior and human dignity. *Journal of Personality and Social Psychology* 1977, 35, 120 – 121.
- KUMPF, M.: Einschätzungen und Konsequenzen der Täuschung von Versuchspersonen in der psychologischen Forschung. In: Kruse, L. & Kumpf, M. (Hrsg.): *Psychologische Grundlagenforschung: Ethik und Recht*. Bern: Huber (im Druck).
- LENNERTZ, E.: *Testwertveränderungen infolge Test-Transparenz*. Unveröffentlichte Dissertation, Universität Bonn, 1969.
- LOVAAS, O. I.: Social desirability ratings of personality variables by Norwegian and American college students. *Journal of Abnormal and Social Psychology* 1958, 57, 124 – 125.
- LÜCK, H. E. & TIMAEUS, E.: Skalen zur Messung Manifesten Angst (MAS) und sozialer Wünschbarkeit (SDS-E und SDS-CM). *Diagnostica* 1969, 15, 134 – 141.
- MAGNUSSON, D. & ENDLER, N. S. (Eds.): *Personality at the crossroads: Current issues in interactional psychology*. Hillsdale, N. J.: Erlbaum, 1977.
- MARLOWE, D. & CROWNE, D. P.: Need for social approval and the operant conditioning of meaningful verbal behavior. *Journal of Consulting Psychology* 1962, 26, 79 – 83.
- MESSICK, S.: *The psychology of acquiescence: An interpretation of research evidence*. Princeton, N. J.: Educational Testing Service, 1966.
- MÉTRAUX, A.: Die Verschränkung methodologischer und ethischer Handlungsregeln. Eine Fallstudie zur Täuschungsproblematik in der Sozialpsychologie. In: Kruse, L. & Kumpf, M. (Hrsg.): *Psychologische Grundlagenforschung: Ethik und Recht*. Bern: Huber (im Druck).
- MIDDLEMIST, D. R., KNOWLES, E. S. & MATTER, C. F.: Personal space invasions in the lavatory: Suggestive evidence for arousal. *Journal of Personality and Social Psychology* 1976, 33, 541 – 546.
- MILGRAM, S.: Behavioral study of obedience. *Journal of Abnormal and Social Psychology* 1963, 67, 371 – 378.
- MILGRAM, S.: Issues in the study of obedience: A reply to Baumrind. *American Psychologist* 1964, 19, 848 – 852.
- MILTURN, L. & LEWIS, M.: Age differences in peer ratings of socially desirable and socially undesirable behavior. *Psychological Reports* 1968, 23, 783 – 791.
- MISCHEL, W.: *Personality and assessment*. New York: Wiley, 1968.

- MORF, M. E. & JACKSON, D. N.: An analysis of two response styles: True responding and item endorsement. *Educational and Psychological Measurement* 1972, 32, 329–353.
- MUMMENDEY, H. D.: *Das Bogus-Pipeline-Paradigma für ethnische Stereotype: Eine experimentelle Replikation mit vereinfachter Versuchsordnung*. Referat auf der 22. Tagung Experimentell Arbeitender Psychologen in Tübingen 1980.
- MUMMENDEY, H. D. & BOLTEN, H.-G.: Die Veränderung von Social-Desirability-Antworten bei erwarteter Wahrheitskontrolle (Bogus-Pipeline-Paradigma). *Zeitschrift für Differentielle und Diagnostische Psychologie* 1981, 2, 151–156.
- MUMMENDEY, H. D., SCHIEBEL, B., TROSKE, U., HESENER, B. & BOLTEN, H.-G.: *Experimentelle Replikation des Bogus-Pipeline-Effekts für ethnische Stereotype*. Bielefelder Arbeiten zur Sozialpsychologie, Nr. 55, Universität Bielefeld, Dezember 1979.
- NORMAN, R. P.: Level of aspiration and social desirability in chronic schizophrenics. *Journal of Consulting Psychology* 1963, 27, 40–44.
- ORNE, M. T.: On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist* 1962, 17, 776–783.
- OSTROM, T. M.: The bogus pipeline: A new ignis fatuus? *Psychological Bulletin* 1973, 79, 252–259.
- PAGE, M. M.: Effects of demand cues and evaluation apprehension in an attitude change experiment. *Journal of Social Psychology* 1973, 89, 55–62.
- REITZNER, C.: *Analyse von Verfälschungstendenzen im Freiburger Persönlichkeitsinventar (FPI) und im Kurzfragebogen für Problemfälle (KFP-30)*. Unveröffentlichte Zulassungsarbeit, Universität Tübingen, Psychologisches Institut, 1974 (zit. n. FAHRENBURG, SELG & HAPPEL, 1978).
- RORER, L. G.: The great response-style myth. *Psychological Bulletin* 1965, 63, 129–156.
- ROSENBERG, M. J.: When dissonance fails: On eliminating evaluation apprehension from attitude measurement. *Journal of Personality and Social Psychology* 1965, 1, 28–42.
- SCHLENKER, B. R. & FORSYTH, D. R.: On the ethics of psychological research. *Journal of Experimental Social Psychology* 1977, 13, 369–396.
- SCHMIDT, H. D. & VORTHMANN, H. R.: Eine Skala zur Messung der „sozialen Erwünschtheit“. *Diagnostica* 1971, 17, 87–90.
- SCHNEIDER, J. F. & HÜBNER, R.: *Einfluß von Verfälschungsinstruktionen auf die Bearbeitungszeit von Persönlichkeitsfragebogen*. Arbeiten der Fachrichtung Psychologie der Universität des Saarlandes, Nr. 68, Universität Saarbrücken, 1980.
- SCHÖNBACH, P.: Likableness ratings of 100 German personality-trait words corresponding to a subset of Anderson's 555 trait words. *European Journal of Social Psychology* 1972, 2, 327, 334.
- SCHULER, H.: *Ethische Probleme psychologischer Forschung*. Göttingen: Hogrefe, 1980.
- SCHULER, H.: Ethische Probleme des psychologischen Forschungsprozesses. Der Stand der Diskussion. In: Kruse, L. & Kumpf, M. (Hrsg.): *Psychologische Grundlagenforschung: Ethik und Recht*. Bern: Huber (im Druck).
- SCHWITZGEBEL, R. K. & KOLB, D. A.: *Systematische Verhaltensänderung*. Stuttgart: Klett-Cotta, 1978.
- SCOTT, W. A.: Social desirability and individual conceptions of the desirable. *Journal of Abnormal and Social Psychology* 1963, 67, 574–585.
- SIGALL, H. & PAGE, R.: Current stereotypes: A little fading, a little faking. *Journal of Personality and Social Psychology* 1971, 18, 247–255.
- STEPHENS, M. W.: *Motivation revisited*. Paper presented at the Institute for Psychiatric Research, Indiana University Medical Center, Indianapolis, Indiana, March 1962 (zit. n. ALLAMAN, JOYCE & CRANDALL, 1972).
- STOLLAK, G. E.: EPPS performance under social desirability instructions. *Journal of Personality and Social Psychology* 1965, 2, 430–432.
- STOLZ, S. B. & Associates. *Ethical issues in behavior modification (Report of the APA commission)*. San Francisco: Jossey-Bass, 1978.
- TAYLOR, J. A.: A personality scale of manifest anxiety. *Journal of Abnormal & Social Psychology* 1953, 48, 285–290.
- TRIANDIS, H. C.: Cultural influences upon cognitive processes. In: Berkowitz, L. (Ed.): *Advances in experimental social psychology* (Vol. 1), 1964, New York: Academic Press, 1964.
- TUCKER, L. R. & MESSICK, S.: An individual differences model for multidimensional scaling. *Psychometrika* 1963, 28, 333–367.
- VELICER, W. F. & WEINER, B. J.: Effects of sophistication and faking sets on the Eysenck Personality Inventory. *Psychological Reports* 1975, 37, 71–73.
- VOYCE, C. D. & JACKSON, D. N.: An evaluation of a threshold theory for personality assessment. *Educational and Psychological Measurement* 1977, 37, 383–408.
- WEST, S. G. & GUNN, S. P.: Some issues of ethics and social psychology. *American Psychologist* 1978, 33, 30–38.
- WICKLUND, R. A.: Objective self-awareness. In: Berkowitz, L. (Ed.): *Advances in experimental social psychology* (Vol. 8). New York: Academic Press, 1975.
- WICKLUND, R. A.: The influence of self-awareness on human behavior. *American Scientist* 1979, 67(2), 187–193. (a)
- WICKLUND, R. A.: Die Aktualisierung von Selbstkonzepten in Handlungsvollzügen. In: Filipp, S.-H. (Hrsg.): *Selbstkonzept-Forschung. Probleme, Befunde, Perspektiven*. Stuttgart: Klett-Cotta, 1979. (b)

- WICKLUND, R. A.: Group contact and self focused attention. In: Paulus, P. B. (Ed.): *Psychology of group influence*. Hillsdale, N. J.: Erlbaum, 1980.
- WIGGINS, J. S.: Interrelations among MMPI measures of dissimulation under standard and social desirability instructions. *Journal of Consulting Psychology* 1959, 23, 419–427.
- WIGGINS, J. S.: *Personality and prediction. Principles of personality assessment*. Reading, Mass.: Addison-Wesley, 1973.
- WIGGINS, J. S.: A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology* 1979, 37, 395–412.
- WIGGINS, N.: Individual viewpoints of social desirability. *Psychological Bulletin* 1966, 66, 68–77.
- WITTICH, U.: *Fragen der Social Desirability und Verfälschungstendenzen im ALNEV-Inventar*. Unveröffentlichte Zulassungsarbeit, Universität Freiburg i. Br., Psychologisches Institut (zit. n. FAHRENBERG, SELG & HAMPEL, 1978).

Prof. Dr. H. D. Mummendey
 Universität Bielefeld
 Postfach 6840
 D 4800 Bielefeld 1